

# CDI Monthly Meeting 20160713

## CDI Monthly Meeting - July 13, 2016

The Community for Data Integration (CDI) meetings are held the 2nd Wednesday of each month from 11:00 a.m. to 12:30 p.m. Eastern Time.

### WebEx:

<https://usgs.webex.com/> - Under the Meeting Center tabs, search for meeting name: "Community for Data Integration."

### Audio:

**USGS/DOI Dial In Number:** (703) 648-4848 (for USGS and DOI offices)

**Toll Free Dial In Number:** (855) 547-8255 (for other offices and telecommute locations)

**Conference Code:** 47919# (same for both numbers)

## Webex Recording

Webex recordings are available to CDI Members approximately 24 hours after the completion of the meeting. Please login to view the recording. If you would like to become a member of CDI, please email [cdi@usgs.gov](mailto:cdi@usgs.gov).

## Agenda (in Eastern time)

11:00a Scientist's Challenge: [A Seismogenic Landslide Database: seeking the best way to make a diverse database accessible to others](#) - Kate Allstadt and Brennah McVey, USGS Hazards

11:10a Welcome - Cheryl Morris - Director for Core Science Analytics, Synthesis & Libraries

11:15a **USGS Cloud Hosting Solutions (CHS) Progress and Activities** - Kimberly Scott, USGS and Vickie Backus, USGS

**Presentation:** Slides are available to CDI Members. Please login to download the slides. If you would like to become a member of CDI, please email [cdi@usgs.gov](mailto:cdi@usgs.gov).

### Abstract:

*This presentation discusses a brief overview of cloud computing and introduces the CHS task order award, CHS team and work breakdown structure (WBS), and planning roadmap. It provides updates on USGS Cloud Hosting Solutions (CHS) progress and activities. It discusses the applications current in the CHS cloud as well as a few brief case study examples.*

### **Presentation Q&A**

Megan Hines: Are there limitations on what types of applications or data we can put into the CHS environment?

Kimberly: Right now, the biggest limitation is with regard to security, so we are only able to handle data that is classified as FISMA low or moderate. For any application that has an identified technical requirement that we have not yet deployed, our team works with the customer and USGS/DOI management to make these solutions available.

Rob: Are there geospatial related tech stacks in place in the sandbox (Geoserver, ArcGIS Server)? We'd like to test performance of geospatial web service endpoints hosted on AWS.

*Additional questions and answers have been posted to the CDI Forum: [Discussion: USGS Cloud Hosting Solutions \(CHS\), 7/13/16](#)*

11:45a **CDI FY15 Project Presentation: Implementing Controlled Vocabulary Services in USGS** - Fran Lightsom, USGS; Peter Schweitzer, USGS; Alan Allwardt, USGS

**Presentation:** Slides are available to CDI Members. Please login to download the slides. If you would like to become a member of CDI, please email [cdi@usgs.gov](mailto:cdi@usgs.gov).

## Abstract:

*The CDI Semantic Web Working Group has been working to improve the recall and precision of searches in USGS data catalogs by making controlled vocabularies conveniently useful. Work began with a CDI project in 2014 and 2015 that developed use cases and prototypes. On July 13 we will demonstrate vocabulary services and several metadata tools that use vocabulary services. We will also talk about the vision, objectives, and next steps of the Working Group's [Controlled Vocabulary Manifesto](#).*

## Presentation Q&A

Cassandra: How can this vocabulary technology be applied to entity and attribute fields in metadata?

Peter: The Entity and Attributes section that you are talking about is the cell values in a table of data. You want to use consistent terms in your data, not just in your metadata. You could build a data editor in which you have an interface to the services and then set up a vocabulary that contains values for those fields. Another way to handle this is to put your vocabulary onto a server and run a checking service to see which values in a database do not match what came from the vocabulary server. So, vocabularies should be used in metadata for keywords, but there is no reason they also shouldn't be used in the data. There is benefit to having standardization in the data as well.

Question: Where are we in USGS in terms of CSDGM vs ISO?

Peter: Most people in USGS are using CSDGM. Mostly because ISO is drastically more complicated. My own feeling is that to make more effective use of ISO we need interfaces for ISO. There are a lot of similarities between the standards. It is really just a question of interfaces. The part of metadata that we are talking about here are the keywords and those fields are used in a really similar way in both standards so implementing vocabulary services should be very similar.

Leslie: There is an upcoming CDI presentation on a project that is working on an ISO metadata tool. That should be in the next few months, so stay tuned.

Viv Hutchison: If you read the policy for the data management requirements, in the metadata chapter, it doesn't specify one or the other. You can use either standard. USGS can poise itself to move in the ISO direction. Otherwise the CSDGM standard can be used.

Alan: Also, note that we need to put metadata into the Science Data Catalog, which currently cannot handle ISO.

Viv: We will be working on accepting ISO metadata in the SDC in the future.

Alan: In the Coastal and Marine Geology Program, we have time series data that uses ISO metadata.

Ra'ad: Can she [Fran] elaborate on the need for a business model for the vocabulary? And how will this model be developed?

Fran: Right now the vocabulary services are living on a mineral resources data server that Peter takes care of. We need to make sure that they don't depend on one person or one program. No clue how to go forward with that. Hopefully, someone will join our working group and help us figure that out.

Peter: The services are located in two locations: On mineral site and science topics site on USGS homepage. This is really a question that all of CDI should think about: When you work across the org chart, there are questions about where you put your stuff because it doesn't fit within one of the mission areas. We talk about collaboration across disciplines and mission areas, but when you do this you need to figure out where to put it. We've talked with WRET about how to handle this, but we didn't come to a conclusion about this.

*Additional questions and answers have been posted to the CDI Forum: [Discussion: Implementing Controlled Vocabulary Services in USGS, 7/13/16](#)*

## 12:20p Working Group Reports

- Citizen Science - Sophia Liu and Dave Govoni
  - On Thursday July 7 in Reston, we held the first USGS Mapathon. We had 55 people (USGS) come out for the event. In future, hopefully we can get more people engaged in the mapping side of things. Two use cases were presented for the mapping projects: The National Map Corps and Mapping Critical Minerals. We hope to do this more frequently with other science centers. If interested contact Sophia Liu ([sophialiu@usgs.gov](mailto:sophialiu@usgs.gov)). For more information on the Mapathon visit <http://bit.ly/USGSMapathon>.
- Communication - JC Nelson and Marcia McNiff
  - The Communication Working Group met yesterday, Tuesday, July 12. We walked through the CDI communication plan section by section and had a great discussion. We will continue working on that in coming weeks. Next we are putting together a PowerPoint to use in CDI presentations. So far, we have 8 people in the working group and invite and encourage others to join.
- Data Management - Heather Henkel and Viv Hutchison
  - The Data Management Working Group met on Monday, July 11. The group talked about [CSASL Workbench Phase I](#), which is designed to help people walk through the five major steps in releasing data. Also, JC Nelson gave an update on the development of USGS data sharing agreement templates, which provide language to use when working with external collaborators around sharing data. Visit the DMWG notes page for more information: [https://my.usgs.gov/confluence/display/cdi/DMWG+Meeting+2016\\_07\\_11](https://my.usgs.gov/confluence/display/cdi/DMWG+Meeting+2016_07_11).
- Earth-Science Themes - Roland Viger
- Semantic Web - Fran Lightsom
  - Tomorrow (July 14) at noon MT the Semantic Web Working Group will have a meeting. We will be talking about all the interesting questions that arose today. We have also been working on a puzzle about how to put semantic web triples data online for the USGS.
- Tech Stack - Rich Signell
  - Meeting today for a Tech Dive.

- Connected Devices - Tim Kern and Lance Everette
  - The June Connected Devices meeting was great. We had an overview of new proposed climate land use app for fisheries data and a follow up presentation about mobile and desktop development frameworks. The group will meet again on July 28th. We will have a follow up to the July 26th BitBucket training and discuss tools and how we implement them for mobile and software development.

12:30p Adjourn

## Attendees

A WebEx Participant Report is available to CDI Members. Please login to download the report. If you would like to become a member of CDI, please email [cdi@usgs.gov](mailto:cdi@usgs.gov).